Kommunikationsstörungen
Berichte aus Phoniatrie und Pädaudiologie
Herausgeber : M. Döllinger
Begründet 1996 von U. Eysholdt

Patrick Schlegel

Assessment of clinical voice parameters and parameter reduction using supervised learning approaches

SHAKER
VERLAG

Assessment of clinical voice parameters and parameter reduction using supervised learning approaches

Beurteilung von klinischen Stimmparametern und Parameterreduzierung unter Verwendung von überwachtem Lernen

Der Technischen Fakultät
der Friedrich-Alexander-Universität
Erlangen-Nürnberg

zur
Erlangung des Doktorgrades
DOKTOR-INGENIEUR

vorgelegt von
Patrick Schlegel
aus Nürnberg

Als Dissertation genehmigt
von der Technischen Fakultät
der Friedrich-Alexander-Universität
Erlangen-Nürnberg

Tag der mündlichen Prüfung:	30. April 2020
Vorsitzender des Promotionsorgans:	Prof. Dr.-Ing. habil. Andreas Paul Fröba
Gutachter:	Prof. Dr.-Ing. Michael Döllinger
	PD. Dr.-Ing. Thomas Wittenberg

Kommunikationsstörungen - Berichte aus Phoniatrie und Pädaudiologie

Band 29

Patrick Schlegel

Assessment of clinical voice parameters and parameter reduction using supervised learning approaches

D 29 (Diss. Universität Erlangen-Nürnberg)

Shaker Verlag
Düren 2020

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: Erlangen-Nürnberg, Univ., Diss., 2020

Printed in Germany.

ISBN 978-3-8440-7435-2
ISSN 1436-1175

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren
Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9
Internet: www.shaker.de • e-mail: info@shaker.de

Danksagung

Mit dem Schreiben dieser Danksagung schließe ich gerade die etwas mehr als zweimonatige Arbeit an dieser Monographie ab, der zum aktuellen Zeitpunkt wiederum fast drei Jahre an der Phoniatrie am Universitätsklinikum in Erlangen vorausgegangen sind. Ich bin kein besonders kommunikativer oder geselliger Mensch und werde euch das nicht oft gesagt haben, aber es war eine schöne Zeit. Aus diesem Grund möchte ich jetzt diese Gelegenheit nutzen um euch meinen Dank auszusprechen.

Zuerst danke ich natürlich meinem Doktorvater Prof. Dr. Michael Döllinger, der mir immer zur Seite stand und mir, der aus einem Nicht-Akademiker Umfeld stammt, unter anderem geholfen hat meinen wissenschaftlichen Schreibstil zu entwickeln. Ohne ihn wäre diese Arbeit vermutlich dick wie ein Telefonbuch und genauso unmöglich zu lesen. Darüber hinaus konnte ich mich auch immer an ihn wenden, wenn ich Fragen hatte und ich weiß, dass ich es vor allem seiner Hilfe zu verdanken habe, dass ich nach meiner Promotion die Möglichkeit habe für eine Post-Doc-Stelle nach Amerika zu gehen. Kurz: Ich hätte mir eigentlich keinen besseren Doktorvater wünschen können.

Ich danke auch meinen Kollegen, die alle auf ihre Weise zu dieser Arbeit beigetragen haben: Ich danke Dr. Pablo Gómez, der den Großteil meiner Promotionszeit mit mir im Büro saß und von dem ich viel über die Kunst des guten und effizienten Programmierens gelernt habe. Ich danke Dr. Andreas Kist, mit dem ich stets einen regen fachlichen Austausch insbesondere zum Thema Machine Learning hatte und der immer ein passendes Fachbuch zu dem Thema, das mich gerade interessierte, parat hatte. Ich danke Dr. Marion Semmler, die an vielen meiner Paper direkt mitgewirkt hat und die mir beigebracht hat, dass selbst eine perfekte Abbildung immer noch etwas perfekter sein kann. Ich danke Dr. Stefan Kniesburges, der ebenfalls viel zu Papern und Konferenzpostern beigetragen hat und der selbst beim Kajak-fahren ein hervorragender Kollege war. Ich danke Sebastian Falk, der gute Laune abstrahlt wie ein gute-Laune-Kernreaktor bei einer gute-Laune-Kernschmelze.

Ich danke Dr. Stephan Dürr, der mich immer auf höchstem fachlichen Niveau endoskopiert hat, wenn ich mit neuen Aufnahmen etwas testen wollte, und der ebenfalls viele meiner Fragen zum klinschen Ablauf und dergleichen beantwortet hat. Ich danke meiner Zweit-Chefin Dr. Anne Schützenberger, für die stets gut gelaunte Antwort auf meine Fragen und dafür, dass sie sich trotz eines überfüllten Terminkalenders immer die Zeit dazu genommen hat, auch wenn es mal etwas länger dauerte. Ich danke Bern-

hard Jakubaß für diverse interessante Gespräche und Fachsimpeleien. Ebenso danke ich meinen Kollegen und ex-Kollegen Gregor Peters (dem "Mukus Man"), Reinhard Veltrup, Sahar Fattoum, Dr. Hossein Sadeghi und Dr. Olaf Wendler für die schöne Zeit.

Ich danke insbesondere all den verschiedenen Ärzten und Logopäden die all die Daten gesammelt haben mit denen ich hier gearbeitet habe und die alle, ohne eine einzige Ausnahme, stets hilfsbereit waren, wenn ich eine Frage an sie hatte oder wieder einmal die zehntausenste Version eines Fragebogens benötigte. Ich danke den Co-Autoren meiner Paper für deren wertvollen Input. Neben den bisher genannten sind das noch insbesondere Prof. Dr. Michael Stingl und Dr. Melda Kunduk die mir mit mathematischer bzw. klinischer Expertise geholfen haben. Außerdem möchte ich mich an dieser Stelle bei allen Leuten bedanken, die ich, (wie könnte es anders sein), vergessen habe aufzuzählen.

Zuletzt danke ich meiner Mutter und meinem Vater, sowie meiner Schwester für die Unterstützung während der Promotionszeit und dafür, dass sie mich letztendlich auf dem Weg geführt haben, der mich zu dem macht, was ich heute bin.

Abstract

Over the years, new measurement techniques and procedures for data collection have been introduced in voice research. Different signals can be measured and on basis of these signals a large number of parameters can be calculated describing different aspects of voice production and quality.

However, since parameters were not introduced in an organized manner, but are rather "grown naturally", analysis on their interdependencies and possible vulnerabilities towards influencing factors has been neglected. If parameters are mathematically dependent they cover the same information. Therefore only one of a group of redundant parameters is needed for data interpretation. Further, if parameters are vulnerable towards certain influencing factors they may not be applicable in some research settings. Results may be distorted by these influencing factors. Also the large number of parameters in use hinders study comparability. Especially in clinical settings a small number of robust parameters is desired, since in these settings different influencing factors that may affect parameters can usually not be avoided. Therefore in this thesis a large scaled investigation of parameters describing voice characteristics and their interrelations is conducted.

In total 963 subjects were examined (180 healthy females, 469 disordered females, 87 healthy males, 227 disordered males). For these subjects 382 high-speed videos and 967 clinical data sets were collected. As part of high-speed video data sets three types of signals were collected. First: The function of changing area between the vocal folds over time, the Glottal Area Waveform (GAW). Second: The Phonovibrogram (PVG) that represents the vibration pattern of the vocal folds over time. And third: parallel recorded audio signals of sustained /i/ vowels. The clinical data sets consisted out of questionnaires and audio signals of sustained /a/ vowels. On basis of the signals derived from high-speed videos 108 parameters (GAT_0) were calculated, on basis of the clinical data sets another 13 parameters (CP_0) were collected.

The influence of changing sequence length on GAW-based perturbation parameters and the influence of changing spatial camera resolution on GAW parameters was investigated. Further mathematical dependencies between different parameters were explored. Parameters in GAT_0 were then reduced by excluding parameters seen as unreliable or redundant based on these investigations. Afterwards a boosted decision stumps approach was applied to find the subset of CP_0 that best separates healthy subjects and two groups of subjects suffering from functional dysphonia (FD). After excluding further linear dependencies between parameters in GAT_0 a similar approach was then used on high-speed video based parameters.

One parameter was found to be significantly influenced by changing sequence length and four parameters were found to be strongly influenced by changing spatial camera resolution. Another nine parameters were found to be redundant. All these parameters were excluded from further analysis.

The reduction of CP_0 using boosted decision stumps yielded parameter set CP_2 consisting out of four parameters. The exclusion of in total 67 parameters from GAT_0 yielded GAT_1 consisting out of 41 parameters. The further reduction of GAT_1 using a modified version of the boosted decision stumps approach yielded GAT_2 consisting out of 12 parameters. The accuracy for different group separations ranged from 0.693 to 0.963 for CP_2 and from 0.752 to 0.793 for GAT_2.

With this work further and new valuable insights in the nature of different parameters and their interdependencies were provided. Parameters were reduced and optimized parameter sets were presented. Insights gained from this work may help researches choosing only the best and most relevant parameter for future research projects. Further this work laid the groundwork for future, more ambitious projects regarding parameter investigation, such as projects that also take into account organic voice disorders that were not investigated in this project. Furthermore, the foundation was laid for the possible development on parameter-based clinical tools. Such tools could provide valuable assistance in assessment and treatment of voice disorders through automated procedures.

Zusammenfassung

Im Verlauf der Jahre wurden in der Stimmforschung immer wieder neue Messtechniken und Verfahren zur Datenerfassung eingeführt. Verschiedene Signale können gemessen werden und auf der Grundlage dieser Signale kann eine große Anzahl von Parametern berechnet werden, die verschiedene Aspekte der Stimmproduktion und -qualität beschreiben.

Da die Parameter jedoch nicht auf organisierte Weise eingeführt wurden, sondern auf "natürliche Weise gewachsen" sind, wurde die Analyse ihrer gegenseitigen Abhängigkeiten und möglichen Anfälligkeiten gegenüber Einflussfaktoren bisher vernachlässigt. Wenn etwa Parameter mathematisch abhängig sind, bilden beide die gleiche Information ab. Daher wird für die Dateninterpretation nur einer aus einer Gruppe redundanter Parameter benötigt. Außerdem sind Parameter, wenn sie gegenüber bestimmten Einflussfaktoren anfällig sind, unter bestimmten Forschungsbedingungen möglicherweise nicht anwendbar. Ergebnisse könnten durch die diese Einflussfaktoren verfälscht werden. Darüber hinaus schränkt die Vielzahl der sich in Gebrauch befindenden Parameter die Vergleichbarkeit von Untersuchungen ein. Insbesondere in klinischen Settings ist eine geringe Anzahl robuster Parameter gewünscht, da in solchen Settings unterschiedliche Einflussfaktoren, gegenüber denen Parameter anfällig sein können, in der Regel nicht vermieden werden können. Daher wird in dieser Arbeit eine umfassende Untersuchung von Parametern welche Stimmcharakteristiken beschreiben, sowie der Wechselbeziehungen zwischen verschiedenen Paramtern, durchgeführt.

Insgesamt wurden 963 Probanden untersucht (180 gesunde Frauen, 469 pathologische Frauen, 87 gesunde Männer, 227 pathologische Männer). Für diese Probanden wurden 382 High-Speed-Videos und 967 klinische Datensätze gesammelt. Basierend auf den High-Speed-Videos wurden drei Arten von Signalen gemessen. Erstens: Die Funktion die die Fläche des Bereichs zwischen den Stimmlippen über die Zeit beschreibt, die Glottal Area Waveform (GAW). Zweitens: Das Phonovibrogramm (PVG), das das Schwingungsmuster der Stimmlippen über die Zeit wiedergibt. Und drittens: parallel aufgezeichneten Audiosignale von gehaltenen /i/ Vokalen. Die klinischen Datensätze bestanden aus Fragebögen und Audiosignalen von gehaltenen /a/ Vokalen. Auf Basis der von High-Speed-Videos abgeleiteten Signale wurden 108 Parameter (GAT_0) berechnet. Auf Basis der klinischen Datensätze wurden weitere 13 Parameter (CP_0) gesammelt.

Der Einfluss der Änderung der Sequenzlänge auf GAW-basierte Perturbationsparameter und der Einfluss der Änderung der räumlichen Kameraauflösung auf GAW-Parameter wurde untersucht. Außerdem wurden weitere mathematische Abhängigkeiten

zwischen verschiedenen Parametern untersucht. Die Parameter in GAT_0 wurden anschließend reduziert, indem Parameter ausgeschlossen wurden, die aufgrund dieser Untersuchungen als unzuverlässig oder redundant angesehen wurden. Anschließend wurde ein Boosted-Decision-Stumps Ansatz angewendet, um die Teilmenge von CP_0 zu finden, welche gesunde Probanden und zwei Gruppen von Probanden mit funktioneller Dysphonie (FD), am besten trennt. Nachdem weitere lineare Abhängigkeiten zwischen Parametern in GAT_0 ausgeschlossen wurden, wurde ein ähnlicher Ansatz für High-Speed-Video basierte Parameter verwendet.

Ein Parameter wurde durch die Änderung der Sequenzlänge signifikant beeinflusst, und vier Parameter wurden durch die Änderung der räumlichen Kameraauflösung stark beeinflusst. Weitere neun Parameter wurden als redundant befunden. Alle diese Parameter wurden von der weiteren Analyse ausgeschlossen.

Die Reduktion von CP_0 durch einen Boosted-Decision-Stumps Ansatz ergab den Parametersatz CP_2 bestehend aus vier Parametern. Der Ausschluss von insgesamt 67 Parametern aus GAT_0 ergab GAT_1, bestehend aus 41 Parametern. Die weitere Reduktion von GAT_1 unter Verwendung einer modifizierten Version des Boosted-Decision-Stumps-Ansatzes resultierte in GAT_2, bestehend aus 12 Parametern. Die Accuracy für verschiedene Gruppentrennungen lag zwischen 0,693 und 0,963 für CP_2 und zwischen 0,752 und 0,793 für GAT_2.

Mit dieser Arbeit wurden weitere und neue wertvolle Erkenntnisse über die Natur der verschiedenen Parameter und ihre gegenseitigen Abhängigkeiten gewonnen. Parameter wurden reduziert und optimierte Parametersätze wurden bereitgestellt. Die aus dieser Arbeit gewonnenen Erkenntnisse könnten Forschern dabei helfen, nur die besten und relevantesten Parameter für ihre zukünftigen Forschungsprojekte auszuwählen. Darüber hinaus bildete diese Arbeit die Grundlage für künftige, ehrgeizigere Projekte zur Parameteruntersuchung, beispielsweise Projekte welche ebenfalls organische Stimmstörungen berücksichtigen, die in diesem Projekt nicht untersucht wurden. Weiterhin wurde der Grundstein für die mögliche Entwicklung auf parametern basierender klinischer Tools gelegt. Solche Tools könnten mithilfe automatisierter Verfahren eine werteveolle Hilfestellung bei der Beurteilung und Behandlung von Stimmstörungen geben.

Contents